

# N-ALGEBRAICITY OF ZETA FUNCTIONS OF SOFIC-DYCK SHIFTS

MARIE-PIERRE BÉAL AND CĂTĂLIN DIMA

**ABSTRACT.** We prove that the multivariate zeta function of a sofic-Dyck shift is the commutative series of some visibly pushdown language. As a consequence the zeta function of a sofic-Dyck shift is the generating function of a visibly pushdown language and is thus an  $\mathbb{N}$ -algebraic series.

## 1. INTRODUCTION

We consider zeta functions which count the periodic orbits of symbolic dynamical systems. The main result of the present paper is the sharpening of algebraicity of the zeta function of sofic-Dyck shifts obtained in [7]. The  $\mathbb{N}$ -algebraicity was obtained for finite-type-Dyck shifts in [8]. We prove here that it holds for all sofic-Dyck shifts, a larger class of symbolic dynamical systems.

The basic concept we shall use is that of  $\mathbb{N}$ -algebraicity. A series  $S(z)$  in one variable  $z$  with coefficients in  $\mathbb{N}$  is called  $\mathbb{N}$ -algebraic if it is the first component  $(S_1(z))$  of a system of equations  $S_i(z) = P_i(z, S_1(z), \dots, P_r(z))$ , where  $1 \leq i \leq r$  and  $P_i$  are multivariate polynomials with coefficients in  $\mathbb{N}$ . The important condition is that one never performs subtraction. This may be interpreted by the existence of an algorithm which generates the objects whose generating function is the given series. Equivalently, a series is  $\mathbb{N}$ -algebraic if and only if it is the generating series of some unambiguous context-free language. Thus  $\mathbb{N}$ -algebraicity is a kind of combinatorial algebraicity.

It was proved by Reutenauer [21] that the zeta function of a sofic shift is  $\mathbb{N}$ -rational. Sofic shifts are symbolic dynamical systems defined as sets of bi-infinite sequences avoiding a regular set of finite patterns. They are equivalently the sets of labels of bi-infinite paths on a labeled directed graph.  $\mathbb{N}$ -rational series are obtained from polynomials over  $\mathbb{N}$  in the variable  $z$  by applying the following operations: sum, product, star of a series (defined if the series  $S(z)$  has no constant term as the inverse of  $1 - S(z)$ ). Equivalently,

---

*Date:* January 26, 2015.

*Key words and phrases.* sofic-Dyck shifts, sofic shifts, symbolic dynamics, visibly pushdown languages, zeta function, multivariate zeta function, rational series, algebraic series.

This work is supported by the French National Agency (ANR) through "Programme d'Investissements d'Avenir" (Project ACRONYME n°ANR-10-LABX-58) and through the ANR EQINOCs.

a series is  $\mathbb{N}$ -rational if and only if it is the generating series of some rational regular language.

In this paper we consider the class of sofic-Dyck shifts introduced in [7], which contains strictly the class of sofic shifts. Sofic-Dyck shifts are sets of bi-infinite sequences avoiding a visibly pushdown language (or nested-word language) of finite patterns. They are defined as sets of labels of admissible bi-infinite paths of a labeled directed graph where some symbols have to be matched with other symbols. Visibly pushdown languages were defined by Mehlhorn [20] and Alur *et al.* [1, 2]. These languages are unambiguous context-free languages which are higher than regular languages in the Chomsky hierarchy but close to them. Indeed, they share many interesting properties of regular languages like stability by intersection and complementation. These languages are used as models for structured data files like XML files. Sofic-Dyck shifts extend Dyck shifts introduced by Krieger in [16] and Markov-Dyck shifts studied by Inoue, Krieger and Matsumoto [17], [13].

In [7] and [8] we obtained an algebraic expression of the zeta function of sofic-Dyck shifts proving its  $\mathbb{Z}$ -algebraicity in general and its  $\mathbb{N}$ -algebraicity for the more restrictive class of finite-type-Dyck shifts. There is no known criterion for a  $\mathbb{Z}$ -algebraic series with coefficients in  $\mathbb{N}$  to be  $\mathbb{N}$ -algebraic although there are some necessary conditions on the asymptotic behavior of the coefficients (see the Drmota-Lalley-Woods Theorem in [11, VII.6.1] and recent results from Banderier and Drmota in [4, 3]). The computation of the zeta functions of sofic-Dyck shifts in [7] is based on ideas and technique for computing the zeta functions of sofic shifts due to Bowen [10] and also on some technique for computing the zeta functions of Dyck shifts and Markov-Dyck shifts from [14], [17], and [12].

The  $\mathbb{N}$ -algebraicity for all sofic-Dyck shifts is obtained with the following tools. We first propose an encoding of periodic patterns which slightly differs from the one used in [7] and [17]. This new encoding allows one to decompose the zeta function of a sofic-Dyck shift as the product of the zeta functions of two shift-invariant sets of bi-infinite sequences. We then show that each zeta function in the product is actually obtained by substituting some letters with visibly pushdown languages in the multivariate zeta function of some sofic shift. The result follows then from the  $\mathbb{N}$ -rationality of the zeta function of a sofic shift obtained by Reutenauer.

## 2. SOFIC-DYCK SHIFTS

This section provides some background on sofic-Dyck shifts introduced in [7]. Sofic-Dyck shifts generalize Markov-Dyck shifts introduced in [16] and [19] (see also [17]).

**2.1. Shifts.** We introduce below some basic notions of symbolic dynamics. We refer to [18, 15] for an introduction to this theory. Let  $A$  be a finite alphabet. The set of finite sequences or words over  $A$  is denoted by  $A^*$  and

the set of nonempty finite sequences or words over  $A$  is denoted by  $A^+$ . The *shift transformation*  $\sigma$  on  $A^{\mathbb{Z}}$  is defined by

$$\sigma((x_i)_{i \in \mathbb{Z}}) = (x_{i+1})_{i \in \mathbb{Z}},$$

for  $(x_i)_{i \in \mathbb{Z}} \in A^{\mathbb{Z}}$ . A *factor* of a bi-infinite sequence  $x$  is a finite word  $x_i \cdots x_j$  for some  $i, j$ , the factor being the empty word if  $j < i$ .

A *subshift* (or *shift*) of  $A^{\mathbb{Z}}$  is a closed shift-invariant subset of  $A^{\mathbb{Z}}$  equipped with the product of the discrete topology. If  $X$  is a shift, a finite word is *allowed* for  $X$  (or is a *block of*  $X$ ) if it appears as a factor of some bi-infinite sequence of  $X$ . We denote by  $\mathcal{B}(X)$  the set of blocks of  $X$ . Let  $F$  be a set of finite words over the alphabet  $A$ . We denote by  $X_F$  the set of bi-infinite sequences of  $A^{\mathbb{Z}}$  avoiding all words of  $F$ , *i.e.* where no factor belongs to  $F$ . The set  $X_F$  is a shift and any shift is the set of bi-infinite sequences avoiding all words of some set of finite words. When  $F$  can be chosen finite (resp. regular), the shift  $X_F$  is called a *shift of finite type* (resp. *sofic*).

Let  $X \subseteq A^{\mathbb{Z}}$  be a shift and  $m, n$  be nonnegative integers. A map  $\Phi : X \rightarrow B^{\mathbb{Z}}$  is called an  $(m, n)$ -*block map* with memory  $m$  and anticipation  $n$  if there exists a function  $\phi : \mathcal{B}_{m+n+1}(X) \rightarrow B$  such that, for all  $x \in X$  and any  $i \in \mathbb{Z}$ ,  $\Phi(x)_i = \phi(x_{i-m} \cdots x_{i-1} x_i x_{i+1} \cdots x_{i+n})$ . A *block map* is a map which is an  $(m, n)$ -block map for some nonnegative integers  $m, n$ .

A *conjugacy* is a bijective block map from  $X$  to  $Y$ . A property of subshifts which is invariant by conjugacies is called a *conjugacy invariant*.

**2.2. Sofic-Dyck shifts and Dyck automata.** We consider an alphabet  $A$  which is a disjoint union of three finite sets of letters, the set  $A_c$  of *call letters*, the set  $A_r$  of *return letters*, and the set  $A_i$  of *internal letters*. The set  $A = A_c \sqcup A_r \sqcup A_i$  is called a *pushdown alphabet*.

We denote by  $\text{MR}(A)$  the set of all finite words over  $A$  where every return symbol is matched with a call symbol, *i.e.*  $u \in \text{MR}(A)$  if for every prefix  $u'$  of  $u$ , the number of call symbols of  $u'$  is greater than or equal to the number of return symbols of  $u'$ . These words are called *matched-return*. Similarly,  $\text{MC}(A)$  denotes the set of all words where every call symbol is matched with a return symbol, *i.e.*  $u \in \text{MC}(A)$  if for every suffix  $u'$  of  $u$ , the number of return symbols of  $u'$  is greater than or equal to the number of call symbols of  $u'$ . These words are called *matched-call*.

We say that a word is a *Dyck word* if it belongs to the intersection of  $\text{MC}(A)$  and  $\text{MR}(A)$ . Dyck words are well-parenthesized or well-formed words. Note that the empty word and all words over  $A_i$  are Dyck words. The set of Dyck words over  $A$  is denoted by  $\text{Dyck}(A)$ .

A (finite) *Dyck automaton*  $\mathcal{A}$  over  $A$  is a pair  $(\mathcal{G}, M)$  of an automaton (or a directed labeled graph)  $\mathcal{G} = (Q, E, A)$  over  $A$  where  $Q$  is the finite set of states,  $E \subseteq Q \times A \times Q$  is the set of edges, and with a set  $M$  of pairs of edges  $((p, a, q), (r, b, s))$  such that  $a \in A_c$  and  $b \in A_r$ . The edges labeled by call letters (resp. return, internal) letters are also called *call* (resp. *return*, *internal*) *edges* and are denoted by  $E_c$  (resp.  $E_r$ ,  $E_i$ ). The set  $M$  is called

the set of *matched edges*. If  $e$  is an edge we denote by  $s(e)$  its starting state and by  $t(e)$  its target state.

A finite path  $\pi$  of  $\mathcal{A}$  is said to be an *admissible path* if for any factor  $(p, a, q) \cdot \pi_1 \cdot (r, b, s)$  of  $\pi$  with  $a \in A_c$ ,  $b \in A_r$  and the label of  $\pi_1$  being a Dyck word on  $A$ ,  $((p, a, q), (r, b, s))$  is a matched pair. Hence any path of length zero is admissible and factors of finite admissible paths are admissible. A bi-infinite path is *admissible* if all its finite factors are admissible.

The *sofic-Dyck shift presented* by  $\mathcal{A}$  is the set of labels of bi-infinite admissible paths of  $\mathcal{A}$  and  $\mathcal{A}$  is called a *presentation* of the shift.

Note that the label of a finite admissible path may not be a block of the presented shift since a finite admissible path may not be extensible to a bi-infinite admissible path. Nevertheless the sofic-Dyck shift presented by a Dyck automaton is exactly the set of bi-infinite sequences  $x$  such that each finite factor of  $x$  is the label of a finite admissible path (see [7]).

*Example 1.* Let  $A = A_c \sqcup A_r \sqcup A_i$  with  $A_c = \{a_1, \dots, a_k\}$ ,  $A_r = \{b_1, \dots, b_k\}$  and  $A_i$  is the empty set. The *Dyck shift* of order  $k$  over the alphabet  $A$  is the set of all sequences accepted by the one-state Dyck automaton  $\mathcal{A} = (\mathcal{G}, M)$  containing all loops  $(p, a, p)$  for  $a \in A$ , and where the edge  $(p, a_i, p)$  is matched with the edge  $(p, b_i, p)$  for  $1 \leq i \leq k$ .

A Motzkin shift is the set of bi-infinite sequences presented by the automaton  $A = A_c \sqcup A_r \sqcup A_i$  with  $A_c = \{a_1, \dots, a_k\}$ ,  $A_r = \{b_1, \dots, b_k\}$ , the set  $A_i$  being no more the empty set. A Motzkin shift is represented in the left part of Figure 1. It is shown in [12] that the entropy of this Motzkin shift, with  $k = 2$  and  $A_i = \{i\}$ , is  $\log 4$ . A sofic-Dyck shift  $X$  is presented by the Dyck automaton in the right part of Figure 1. For instance, the bi-infinite sequences  $\dots(( [ i i ] [ ] ) \dots$  and  $\dots)) \dots$  belong to  $X$  while the sequence  $\dots([ i ] [ ] ) \dots$  does not.

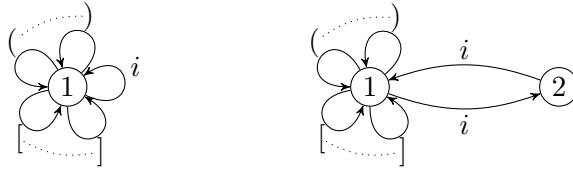


FIGURE 1. A Motzkin shift (on the left) over  $A = A_c \sqcup A_r \sqcup A_i$  with  $A_c = \{([, ]), A_r = \{), \}$  and  $A_i = \{i\}$ . A sofic-Dyck shift (on the right) over the same tri-partitioned alphabet. Matched edges are linked with a dotted line.

Let  $L$  be a language of finite words. A Dyck automaton is *L-deterministic* (resp. *L-codeterministic*) if there is at most one admissible path starting at (resp. ending in) a given state and with a given label in  $L$ .

It is proved in [7] that sofic-Dyck shifts are the shifts  $X_F$  where  $F$  is a visibly pushdown language.

Let  $A$  and  $B$  be two tri-partitioned alphabets. We say that a block-map  $\Phi : A^{\mathbb{Z}} \rightarrow B^{\mathbb{Z}}$  is *proper* if and only if  $\Phi(x)_i \in A_c$  (resp.  $A_r, A_i$ ) whenever  $x_i \in A_c$  (resp.  $A_r, A_i$ ). Sofic-Dyck shifts are invariant by proper conjugacies.

We call *left reduction* of a Dyck automaton  $\mathcal{A}$  a Dyck automaton presenting the same sofic-Dyck shift as  $\mathcal{A}$  which is obtained with some determinization process from  $\mathcal{A}$  (see [6]). The left reduction of  $\mathcal{A}$  is  $A_c$ -deterministic and Dyck( $A$ )-deterministic. We define similarly the *right reduction* of a Dyck automaton obtained by co-determinization which is  $A_r$ -codeterministic and Dyck( $A$ )-codeterministic.

We say that  $\mathcal{A}$  is *left reduced* (resp. *right reduced*) it is the left (resp. right) reduction of some Dyck automaton. Left and right reduced presentations will be useful to compute the zeta function of the sofic-Dyck shifts.

### 3. ZETA FUNCTIONS

For each language  $L$  of finite words over a finite alphabet  $A$  we define the characteristic series of  $L$  as  $\underline{L} = \sum_{u \in L} u$ .

Denote  $\mathbb{Z}\langle\langle A \rangle\rangle$  the set of noncommutative formal power series over the alphabet  $A$  with coefficients in  $\mathbb{Z}$ . Let  $\mathbb{Z}[[A]]$  be the usual commutative algebra of formal power series in the variables  $a$  in  $A$  and  $\pi : \mathbb{Z}\langle\langle A \rangle\rangle \rightarrow \mathbb{Z}[[A]]$  be the natural homomorphism. Let  $S \in \mathbb{Z}\langle\langle A \rangle\rangle$  be a commutative or noncommutative series. One can write  $S = \sum_{n \geq 0} [S]_n$  where each  $[S]_n$  is the homogeneous part of  $S$  of degree  $n$ . The notation extends to  $(Q \times Q)$ -matrices with coefficients in  $\mathbb{Z}\langle\langle A \rangle\rangle$  or  $\mathbb{Z}[[A]]$  with  $([H]_n)_{pq} = [H_{pq}]_n$ .

The *multivariate zeta function* of  $S$  is the commutative series in  $\mathbb{Z}[[A]]$

$$Z(S) = \exp \sum_{n \geq 1} \frac{[S]_n}{n}.$$

The (*ordinary*) *zeta function* of a language  $L$  is

$$\zeta_L(z) = \exp \sum_{n \geq 1} a_n \frac{z^n}{n},$$

where  $a_n$  is the number of words of lengths  $n$  of  $L$ . Note that  $\zeta_L(z) = \theta(Z(\underline{L}))$ , where  $\theta : \mathbb{Z}[[A]] \rightarrow \mathbb{Z}[[z]]$  is the homomorphism such that  $\theta(a) = z$  for any letter  $a \in A$ .

Call *periodic pattern* of  $X$  a word  $u$  such that the bi-infinite concatenation of  $u$  belongs to  $X$  and denote  $P(X)$  the set of periodic patterns of  $X$ . These definitions are extended to  $\sigma$ -invariant sets of bi-infinite sequences which may not be shifts (*i.e.* which may not be closed subsets of sequences).

The *multivariate zeta function*  $Z(X)$  of the shift (resp. *zeta function*  $\zeta_X(z)$  of the shift)  $X$  is defined as the multivariate zeta function (resp. zeta function) of the commutative series of its set of periodic patterns. Hence

$$\zeta_X(z) = \exp \sum_{n \geq 1} p_n \frac{z^n}{n},$$

where  $p_n$  the number of points of  $X$  of period  $n$ , *i.e.* of points  $x$  such that  $\sigma^n(x) = x$ .

#### 4. ENCODING OF PERIODIC POINTS

We say that a Dyck word  $w$  over  $A$  is *prime* if it is nonempty and any Dyck word prefix of  $w$  is  $w$  or the empty word.

Let  $\mathcal{A}$  be a Dyck automaton over a tri-partitioned alphabet  $A$ . We define the following matrices where  $Q$  is the set of states of  $\mathcal{A}$ .

- $C = (C_{pq})$ , where  $C_{pq}$  is the set of prime Dyck words labeling an admissible path from  $p$  to  $q$  in  $\mathcal{A}$ .
- $M_c = (M_{c,pq})$ , (resp.  $M_r$ ) where  $M_{c,pq}$  is the sum of call (resp. return) letters labeling an edge from  $p$  to  $q$  in  $\mathcal{A}$ .

Let  $H$  be one of the matrices  $C, C^*M_c, M_r+C, M_rC^*, M_c+C$ . We call  $H$ -*path* a path  $(p_i, c_i, p_{i+1})_{i \in I}$  in  $\mathcal{A}$ , where  $I$  is  $\mathbb{Z}$  or an interval and  $c_i \in H_{p_i p_{i+1}}$ . Note that an  $H$ -path is admissible. With a slight abuse of notation, we will say that a finite word  $u$  belongs to  $H$  if  $u$  belongs to  $H_{pq}$  for some states  $p, q$ . We denote by  $X_H$  be the  $\sigma$ -invariant set containing all orbits of points labeling a bi-infinite  $H$ -path of  $\mathcal{A}$ .

Let  $P(X)$  denotes the set of periodic patterns of a  $\sigma$ -invariant set  $X$ .

**Proposition 1.** *Let  $X$  be a the sofic-Dyck shift accepted by a Dyck automaton  $\mathcal{A}$ . We have*

$$P(X) = P(X_{C^*M_c}) \sqcup P(X_{M_r+C}),$$

where  $\sqcup$  denotes a disjoint union.

If  $u$  is a word over  $A$ , we denote by  $\text{bal}(u)$  the *balance* of  $u$ , *i.e.* the difference between the number of letters of  $u$  in  $A_c$  and the number of letters of  $u$  in  $A_r$ .

We say that two finite words  $u, v$  are *conjugate* if there are words  $z, t$  such that  $u = zt$  and  $v = tz$ .

*Proof.* Let us assume that a point  $x \in X$  is equal to  $u^\infty = \cdots uu \cdot uu \cdots$ , where  $u = u_0 u_1 \cdots u_{n-1}$ , with  $u_i \in A$  and  $n \geq 1$ . We consider the following three cases.

- If  $\text{bal}(u) = 0$ , then  $u$  is conjugate to a Dyck word and thus  $x$  is a periodic pattern of  $X_C$  and thus of  $X_{M_r+C}$ .
- If  $\text{bal}(u) > 0$ , then  $u$  is conjugate to a word  $v$  such that  $\text{bal}(v_0 \dots v_i) \geq 0$  for any  $0 \leq i \leq n-1$ . If  $v \in A_c^+$ , then  $x$  is a periodic point of  $X_{M_c}$  and thus of  $X_{C^*M_c}$ . If  $v \notin A_c^+$ , there are two indices  $0 \leq m_1 < m_2 \leq n-1$  such that  $\text{bal}(v_0 \dots v_{m_1}) = \text{bal}(v_0 \dots v_{m_2})$ . Let  $m_1, m_2$  be two such indices such that we have moreover  $\text{bal}(v_0 \dots v_{m_1}) = \text{bal}(v_0 \dots v_{m_2})$  minimal. Let  $w = v_{m_1} \cdots v_{n-1} v_0 \cdots v_{m_1-1}$ . The word  $w$  is again a conjugate of  $v$  and  $w$  and starts with a prime Dyck word.

Let  $j_1$  be the largest integer less than or equal to  $n-1$  such that  $w_0 \cdots w_{j_1}$  has a suffix which is a prime Dyck word and  $i_1$  be the smallest integer such that  $w_{i_1} \cdots w_{j_1}$  is a Dyck word. Then  $w_{i_1} \cdots w_{n-1}$  is the label of an  $(C^*M_c)$ -path.

We define indices  $i_2, j_2$  similarly for the word  $w_0 \cdots w_{i_1-1}$  and thus iteratively decompose  $w$  into a product of  $(C^*M_c)$ -words. It follows that  $x$  belongs to  $X_{C^*M_c}$ .

- If  $\text{bal}(u) < 0$ , we perform a similar decomposition by reading words from right to left. We denote by  $\tilde{u}$  the word  $u_{n-1}u_{n-2} \cdots u_0$ . By exchanging the role of played by call and return edges, we have  $\text{bal}(\tilde{u}) > 0$  and thus  $\tilde{u}$  is conjugate to word labeling a  $(\tilde{C}^*M_r)$ -path, where  $\tilde{C} = \{\tilde{c} \mid c \in C\}$ . Hence  $u$  is conjugate to a word labeling an  $(M_rC^*)$ -path. We get that  $x$  belongs to  $X_{M_r+C}$ .

Let us finally show that  $P(X_{C^*M_c}) \cap P(X_{M_r+C}) = \emptyset$ . Let us assume the contrary. Then there are nonempty conjugate words  $u, v$  such that  $u$  is the label of an  $(C^*M_c)$ -path and  $v$  is the label of an  $(M_r + C)$ -path. Since  $\text{bal}(u) = \text{bal}(v)$  and  $\text{bal}(u) > 0$  and  $\text{bal}(v) \leq 0$ , we get a contradiction.  $\square$

As a consequence, we obtain the following expression of the zeta function of a sofic-Dyck shift.

**Proposition 2.** *Let  $X$  be a sofic-Dyck shift presented by a Dyck automaton  $\mathcal{A}$  and  $C, M_r, M_c$  defined as above from  $\mathcal{A}$ . The multivariate zeta function of  $X$  is*

$$Z(X) = Z(X_{C^*M_c})Z(X_{M_r+C}).$$

*Proof.* It is a direct consequence of the definitions and Proposition 1.  $\square$

## 5. COMPUTATION OF THE ZETA FUNCTION

In this section,  $\mathcal{A}$  denotes a Dyck automaton over  $A$  and  $H$  is one of the matrices  $C, C^*M_c, M_r + C, M_rC^*, M_c + C$  defined from  $\mathcal{A}$ .

We say that the matrix  $(H_{pq})_{p,q \in Q}$ , where each  $H_{pq}$  is a set of nonempty words over  $A$  is *circular* if for all  $n, m \geq 1$  and  $x_1 \in H_{p_0, p_1}, x_2 \in H_{p_1, p_2}, \dots, x_n \in H_{p_{n-1}, p_0}, y_1 \in H_{q_0, q_1}, y_2 \in H_{q_1, q_2}, \dots, y_m \in H_{q_{m-1}, q_0}$  and  $p \in A^*$  and  $s \in A^+$ , the equalities

$$(1) \quad sx_2x_3 \cdots x_np = y_1y_2 \cdots y_m,$$

$$(2) \quad x_1 = ps$$

implies

$$n = m \quad p = \varepsilon \quad \text{and} \quad x_i = y_i \quad (1 \leq i \leq n).$$

This notion extends the classical notion of circular codes (see for instance [9]).

**Proposition 3.** *Let  $\mathcal{A}$  be a Dyck automaton. The matrices  $C^*M_c$  and  $M_r + C$  defined from  $\mathcal{A}$  are circular matrices.*

*Proof.* We first show that  $M_r + C$  is circular.

Let us suppose that Equations 1 and 2 imply  $n = m$  and  $x_i = y_i$  for  $n+m < N$ . Assume that Equations 1 and 2 hold for some  $n, m$  with  $n+m = N$ .

If  $x_1 \in A_r$ , then  $s \in A_r$  and  $p = \varepsilon$ . Since  $s$  and  $y_1$  share a nonempty prefix, we have  $y_1 = x = x_1$ . By induction hypothesis, we obtain that  $n = m$  and  $x_i = y_i$ .

If  $s = x_1 \in C$  then  $p = \varepsilon$  and this case is similar as the previous one.

If  $x_1 \in C$  and  $s \neq x_1$ , then  $p \neq \varepsilon$ . As  $x_1 = ps$ , we get  $\text{bal}(p) > 0$ . This contradicts the fact that  $p$  is a suffix of some  $y_j y_{j+1} \cdots y_m$ .

We now show that  $C^* M_c$  is circular.

Let us assume that  $s \neq x_1$ . Since  $s$  is a prefix of some  $y_1 y_2 \cdots y_j$  and a suffix of  $x_1$ , we have  $s \in C^* M_c$  and  $p \in C^*$ . As  $p \neq \varepsilon$ ,  $p \in C^+$  and  $p \notin A^* A_c$ . This contradicts the fact that  $p$  is a suffix of some  $y_k y_{k+1} \cdots y_m$ . Hence  $s = x_1$  and  $p = \varepsilon$ . We get  $y_1 = x_1$  since  $C^* M_c$  is a prefix-free set. By iteration of this process, we get  $n = m$  and  $x_i = y_i$ .  $\square$

We say that  $\mathcal{A}$  is *H-deterministic* if and only if for any two  $H$ -paths sharing the same start and label, are equal.

**Proposition 4.** *A left-reduced Dyck automaton is H-deterministic for  $H = C^* M_c$ .*

*Proof.* Let us show that  $\mathcal{A}$  is  $(C^* M_c)$ -deterministic. Let  $w$  be the label of an  $(C^* M_c)$ -path. It has a unique decomposition  $x_1 \dots x_k$  with  $x_i \in (A_c + \text{Dyck}(A))$  since  $C^* M_c$  is circular and thus a code. The result follows from the fact that a left-reduced Dyck automaton is  $M_c$ -deterministic and  $C$ -deterministic (see [6]).  $\square$

One proves a similar property for left-reduced presentations.

**Proposition 5.** *A right-reduced Dyck automaton is H-codeterministic for  $H = M_r + C$ .*

We get from Proposition 2 that the computation of the multivariate zeta function of a sofic-Dyck shift  $X$  presented by a Dyck automaton  $\mathcal{A}$  reduces to the computation of the multivariate zeta function of  $X_{C^* M_c}$  and  $X_{C+M_r}$ .

**Proposition 6.** *Let  $\mathcal{A}$  (resp.  $\mathcal{A}'$ ) be a left-reduced (resp. a right-reduced) Dyck automaton presenting  $X$ . Let  $C, M_r, M_c$  (resp.  $C', M'_r, M'_c$ ) defined from  $\mathcal{A}$  (resp. from  $\mathcal{A}'$ ) as above. The multivariate zeta function of  $X$  has the following expression.*

$$Z(X) = \prod_{\ell=1}^{|Q|} \det(I - (C^* M_c)_{\otimes \ell})^{(-1)^\ell} \prod_{\ell=1}^{|Q'|} \det(I - (C' + M'_r)_{\otimes \ell})^{(-1)^\ell}.$$

*Proof.* From Proposition 2, we get that the multivariate zeta function of  $X$  is  $Z(X) = Z(X_{C^* M_c}) Z(X_{C+M_r})$ . From [6, Proposition 17] and since  $C^* M_c$  and  $C + M_r$  are circular matrices, we have

$$Z(X_{C^* M_c}) = \prod_{\ell=1}^{|Q|} \det(I - (C^* M_c)_{\otimes \ell})^{(-1)^\ell},$$



$$Z(\mathbf{X}_{C+M_r}) = \prod_{\ell=1}^{|Q'|} \det(I - (C' + M'_r)_{\otimes \ell})^{(-1)^\ell},$$

hence the proposition.  $\square$

*Example 2.* Let  $X$  be the sofic-Dyck presented by the Dyck automaton  $\mathcal{A}$  pictured on the left part of Figure 2, where  $A = (\{a, a'\}, \{b, b'\}, \{i\})$ . The Dyck automaton  $\mathcal{A}_{\otimes 1}$  is the same as  $\mathcal{A}$ . The Dyck automaton  $\mathcal{A}_{\otimes 2}$  is pictured on the right part of Figure 2. Let

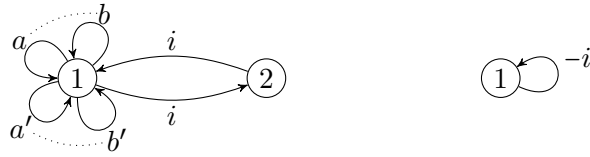


FIGURE 2. A Dyck automaton  $\mathcal{A}$  (on the left) over  $A = A_c \sqcup A_r \sqcup A_i$  with  $A_c = \{a, a'\}$ ,  $A_r = \{b, b'\}$  and  $A_i = \{i\}$  and the Dyck automaton  $\mathcal{A}_{\otimes 2}$  (on the right). Matched edges are linked with a dotted line.

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \quad C_{\otimes 2} = [C_{(1,2),(1,2)}].$$

We have  $C_{11} = aD_{11}b + a'D_{11}b'$ ,  $C_{22} = 0$ ,  $C_{12} = i$ ,  $C_{21} = i$ , with  $D_{11} = aD_{11}bD_{11} + a'D_{11}b'D_{11} + iD_{11} + \varepsilon$ . We also have  $C_{(1,2),(1,2)} = -i$ .

We have

$$\begin{aligned} (C^* M_c)_{\otimes 1} &= C^* M_c = \begin{bmatrix} C_{11} & i \\ i & 0 \end{bmatrix}^* \begin{bmatrix} a + a' & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} (C_{11} + i^2)^* & (C_{11} + i^2)^* i \\ i(C_{11} + i^2)^* & i(C_{11} + i^2)^* i \end{bmatrix} \begin{bmatrix} a + a' & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} (C_{11} + i^2)^*(a + a') & 0 \\ i(C_{11} + i^2)^*(a + a') & 0 \end{bmatrix}, \\ (C^* M_c)_{\otimes 2} &= [0], \\ (C + M_r)_{\otimes 1} &= (C + M_r) = \begin{bmatrix} C_{11} + b + b' & i \\ i & 0 \end{bmatrix}, \\ (C + M_r)_{\otimes 2} &= [-i]. \end{aligned}$$

We get

$$Z(X) = \frac{(1 + i)}{(1 - (C_{11} + i^2)^*(a + a'))(1 - (C_{11} + i^2 + b + b'))}.$$

## 6. N-ALGEBRAICITY

Let  $\mathcal{A}$  be a Dyck automaton and  $H$  a  $(Q \times Q)$ -matrix with coefficients in  $\mathbb{Z}[[A]]$ , where  $Q$  is the set of states of  $\mathcal{A}$ . We shall construct some generic automaton whose letter action on the states of  $\mathcal{A}$  simulate the action of the words in  $H$ .

For each subset  $\mathcal{S}$  of the set of all pairs of states  $\mathcal{P}$  of  $\mathcal{A}$ , we denote by  $V_{H,\mathcal{S}}$  the set of words  $u$  such that  $u \in H_{pq}$  if and only if  $(p, q) \in \mathcal{S}$ . Note that the sets  $V_{H,\mathcal{S}}$  are disjoint for distinct sets  $\mathcal{S}$ . We denote by  $A_H$  the finite alphabet  $\{a_{H,\mathcal{S}} \mid \mathcal{S} \subseteq \mathcal{P}\}$ . Each letter  $a_{H,\mathcal{S}}$  represents the language  $V_{H,\mathcal{S}}$ . We denote by  $\mathcal{G}_H$  the directed graph labeled in  $A_H$  and  $(p, b, q)$  is an edge of  $\mathcal{G}_H$  if and only  $b = a_{H,\mathcal{S}}$  with  $(p, q) \in \mathcal{S}$ . The labeled graph  $\mathcal{G}_H$  presents a sofic shift  $S_H$  over  $A_H$ .

**Proposition 7.** *For any subset  $\mathcal{S}$  of the set of all pairs of states of  $\mathcal{A}$ ,  $V_{H,\mathcal{S}}$  is a visibly pushdown language.*

*Proof.* By definition,  $V_{H,\mathcal{S}} = \cap_{p,q \in \mathcal{S}} H_{pq} \setminus \cup_{p,q \in \mathcal{P} \setminus \mathcal{S}} H_{pq}$ . Since the union and the intersection of two visibly pushdown languages over  $A$  are visibly pushdown languages over  $A$ , the set  $V_{H,\mathcal{S}}$  is a visibly pushdown language over  $A$ .  $\square$

We denote by  $G_H$  the transition matrix of  $\mathcal{G}_H$ . Hence

$$G_{H,pq} = \sqcup_{(p,q) \in \mathcal{S}} a_{H,\mathcal{S}},$$

where  $\sqcup$  denotes a disjoint union.

We denote by  $\theta_H: \mathbb{Z}[[A_H]] \rightarrow \mathbb{Z}[[A]]$  is the homomorphism such that  $\theta_H(a_{H,\mathcal{S}}) = V_{H,\mathcal{S}}$  for any letter  $a_{H,\mathcal{S}} \in A_H$ . The two following propositions are direct consequences of the definitions.

**Proposition 8.** *For any  $1 \leq \ell \leq |Q|$ , we have,*

$$\begin{aligned} H &= \theta_H(G_H), \\ H_{\otimes \ell} &= \theta_H(G_H)_{\otimes \ell}. \end{aligned}$$

**Proposition 9.** *For any  $1 \leq \ell \leq |Q|$ , we have*

$$G_{H_{\otimes \ell}} = (G_H)_{\otimes \ell}.$$

As a consequence, we get the following proposition.

**Proposition 10.**

$$Z(X) = \theta_H(Z(S_H))\theta_{H'}(Z(S_{H'})).$$

*Proof.* From Proposition 2, we have

$$Z(X) = \prod_{\ell=1}^{|Q|} \det(I - H_{\otimes \ell})^{(-1)^\ell} \prod_{\ell=1}^{|Q'|} \det(I - H'_{\otimes \ell})^{(-1)^\ell}.$$

From Propositions 8 and 9, we obtain

$$Z(X) = \prod_{\ell=1}^{|Q|} \det(I - \theta_H(G_H)_{\otimes \ell})^{(-1)^\ell} \prod_{\ell=1}^{|Q'|} \det(I - \theta_{H'}(G'_{H'})_{\otimes \ell})^{(-1)^\ell}.$$

Finally, since  $G_H$  (resp.  $G'_{H'}$ ) is an unambiguous presentation of a sofic shift, we obtain from Bowen's formula of the multivariate zeta function of a sofic shift (see [18], [5]).

$$\begin{aligned} Z(X) &= \theta_H \left( \prod_{\ell=1}^{|Q|} \det(I - (G_H)_{\otimes \ell})^{(-1)^\ell} \right) \theta'_{H'} \left( \prod_{\ell=1}^{|Q'|} \det(I - (G'_{H'})_{\otimes \ell})^{(-1)^\ell} \right), \\ &= \theta_H Z(S_H) \theta_{H'} Z(S_{H'}). \end{aligned}$$

□

We say that a commutative series in  $\mathbb{N}[[A]]$  is an  $\mathbb{N}$ -visibly pushdown series if it is the commutative image of the characteristic series of some visibly pushdown language.

**Theorem 1.** *The multivariate zeta function of a sofic-Dyck shift is computable  $\mathbb{N}$ -visibly pushdown series, i.e. is the commutative series of some visibly pushdown language.*

*Proof.* Let  $X$  be a sofic-Dyck shift. From Proposition 10, the multivariate zeta function of  $X$  is  $Z(X) = \theta_H(Z(S_H))\theta_{H'}(Z(S_{H'}))$  where  $S_H$  (resp.  $S'_{H'}$ ) is a sofic shift over  $A_H$  (resp.  $A_{H'}$ ) and  $\theta_H$  (resp.  $\theta_{H'}$ ) is the substitution replacing each symbol  $a_{H,S}$  of  $A_H$  (resp.  $a_{H',S'}$  of  $A_{H'}$ ) by the visibly pushdown language  $V_{H,S}$  (resp.  $V_{H',S'}$ ). Reutenauer's theorem [21] assures that the multivariate zeta function of a sofic shift is  $\mathbb{N}$ -rational, or equivalently is the commutative series of a rational language. Substituting each letter in a regular expression over  $A_H$  with a visibly pushdown language over  $A$  gives the commutative series of a visibly pushdown language. Hence  $Z(X)$  is a commutative series of a visibly pushdown language as a product of two series of this type. □

We derive the following theorem.

**Theorem 2.** *The zeta function of a sofic-Dyck shift is the generating series of some computable visibly pushdown language. It is thus  $\mathbb{N}$ -algebraic.*

*Proof.* Since a visibly pushdown language is an unambiguous context-free language, the generating series of a visibly pushdown language is  $\mathbb{N}$ -algebraic. □

*Example 3.* We continue Example 2. Let  $\mathcal{A}$  be the Dyck automaton of 2. Let  $H = C^*M_c$  and  $H' = C + M_r$ .

We define the finite-state automata  $\mathcal{G}_H$  over  $A_H = \{a_1, \dots, a_3\}$  where  $a_1 = a_{H,(1,1)}$ ,  $a_2 = a_{H,(1,2)}$ ,  $a_3 = a_{H,(2,1)}$  and  $\mathcal{G}'_{H'}$  over  $A_{H'} = \{b_1, \dots, b_5\}$  where  $b_1 = a_{H',(1,1)}$ ,  $b_2 = a_{H',(1,2)}$ ,  $b_3 = a_{H',(2,1)}$ ,  $b_4 = a_{H',(2,2)}$ ,  $b_5 = a_{H',(1,2),(2,1)}$ .

The automata  $\mathcal{G}_H$  and  $\mathcal{G}_{H'}$  are presentations of two sofic shifts denoted  $S_H$  and  $S_{H'}$  respectively. By definition

$$\begin{aligned} \theta_H(a_1) &= (C_{11} + i^2)^*(a + a'), \\ \theta_H(a_3) &= i(C_{11} + i^2)^*(a + a'), \end{aligned}$$



FIGURE 3. The automaton  $\mathcal{G}_H$  (on the left), over  $A_H = \{a_1, \dots, a_3\}$  where  $H = C^*M_c$ ,  $a_1 = a_{H,(1,1)}$ ,  $a_2 = a_{H,(1,2)}$ ,  $a_3 = a_{H,(2,1)}$  and the automaton  $\mathcal{G}'_{H'}$  (on the right) over  $A_{H'} = \{b_1, \dots, b_5\}$  where  $b_1 = a_{H',(1,1)}$ ,  $b_2 = a_{H',(1,2)}$ ,  $b_3 = a_{H',(2,1)}$ ,  $b_4 = a_{H',(2,2)}$ ,  $b_5 = a_{H',(1,2),(2,1)}$ .

$$\begin{aligned}\theta_{H'}(b_1) &= C_{11} + b + b', \\ \theta_{H'}(b_5) &= i.\end{aligned}$$

N-rational expressions of  $Z(S_H)$  and  $Z(S_{H'})$  are the following.

$$\begin{aligned}Z(S_H) &= \frac{1}{1 - a_1} = a_1^*, \\ Z(S_{H'}) &= \frac{1 + b_5}{1 - b_1 - b_5^2} = (1 + b_5)(b_1 + b_5^2)^*.\end{aligned}$$

It follows that

$$Z(X) = \theta_H \theta'_H ((a_1)^* (1 + b_5) (b_1 + b_5^2)^*),$$

We obtain the following expression of  $Z(X)$  as an N-visibly pushdown series.

$$Z(X) = ((C_{11} + i^2)^* (a + a'))^* (1 + i) (C_{11} + b + b' + i^2)^*,$$

where  $C_{11}$  is the visibly pushdown language of Dyck words labeling paths going from 1 to 1 in  $\mathcal{A}$ .

We obtain

$$\zeta_X(z) = ((C_{11}(z) + z^2)^* (2z))^* (1 + z) (C_{11}(z) + 2z + z^2)^*.$$

The computation of  $C_{11}(z)$  can be performed as follows. We have  $C_{11} = aD_{11}b + a'D_{11}b'$ , with  $D_{11} = aD_{11}bD_{11} + a'D_{11}b'D_{11} + iiD_{11} + \varepsilon$ . It gives

$$2z^2 D_{11}^2(z) - (1 - z^2) D_{11}(z) + 1 = 0$$

Since the coefficient of  $z^0$  in  $D_{11}(z)$  is 1, we get

$$D_{11}(z) = \frac{1 - z^2 - \sqrt{1 - 10z^2 + z^4}}{4z^2}.$$

Hence

$$C_{11}(z) = 2z^2 D_{11}(z) = \frac{1 - z^2 - \sqrt{1 - 10z^2 + z^4}}{2}.$$

## REFERENCES

- [1] R. Alur and P. Madhusudan. Visibly pushdown languages. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 202–211 (electronic), New York, 2004. ACM.
- [2] R. Alur and P. Madhusudan. Adding nesting structure to words. *J. ACM*, 56(3), 2009.
- [3] C. Banderier and M. Drmota. Coefficients of algebraic functions: formulae and asymptotics. *Combinatorics, Probability and Computing*. Special issue dedicated to Philippe Flajolet, to appear.
- [4] C. Banderier and M. Drmota. Coefficients of algebraic functions: formulae and asymptotics. In *DMTCS Proc. AS, 25th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2013)*, pages 1065–1076, 2013.
- [5] M.-P. Béal. Puissance extérieure d’un automate déterministe, application au calcul de la fonction zêta d’un système sofique. *RAIRO Inform. Théor. Appl.*, 29, 1995.
- [6] M.-P. Béal, M. Blockelet, and C. Dima. Sofic-Dyck shifts. *CoRR*, <http://arxiv.org/1305.7413>, 2013. submitted.
- [7] M.-P. Béal, M. Blockelet, and C. Dima. Sofic-Dyck shifts. In *39th International Symposium on Mathematical Foundations of Computer Science MFCS 2014, (MFCS 2014)*, Lecture Notes in Computer Science. Springer-Verlag, 2014.
- [8] M.-P. Béal, M. Blockelet, and C. Dima. Zeta functions of finite-type-Dyck shifts are  $\mathbb{N}$ -algebraic. In *2014 Information Theory and Applications Workshop, ITA 2014*. IEEE, 2014.
- [9] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*, volume 129 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.
- [10] R. Bowen. Symbolic dynamics. In *On axiom A diffeomorphism*, number 35 in CBMS Reg. Conf. American Mathematical Society, 1978.
- [11] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009.
- [12] K. Inoue. The zeta function, periodic points and entropies of the Motzkin shift. *CoRR*, 2006. [math/0602100](http://arxiv.org/0602100).
- [13] K. Inoue and W. Krieger. Subshifts from sofic shifts and Dyck shifts, zeta functions and topological entropy. *CoRR*, 2010. [abs/1001.1839](http://arxiv.org/abs/1001.1839).
- [14] G. Keller. Circular codes, loop counting, and zeta-functions. *J. Combin. Theory Ser. A*, 56(1):75–83, 1991.
- [15] B. P. Kitchens. *Symbolic Dynamics*. Universitext. Springer-Verlag, Berlin, 1998. One-sided, two-sided and countable state Markov shifts.
- [16] W. Krieger. On the uniqueness of the equilibrium state. *Math. Systems Theory*, 8(2):97–104, 1974/75.
- [17] W. Krieger and K. Matsumoto. Zeta functions and topological entropy of the Markov-Dyck shifts. *Münster J. Math.*, 4:171–183, 2011.
- [18] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
- [19] K. Matsumoto.  $C^*$ -algebras arising from Dyck systems of topological Markov chains. *Math. Scand.*, 109(1):31–54, 2011.
- [20] K. Mehlhorn. Pebbling mountain ranges and its application of dcfl-recognition. In *Automata, Languages and Programming, 7th Colloquium, Noordwijkerhout, The Netherlands, July 14-18, 1980, Proceedings*, volume 85 of *Lecture Notes in Computer Science*, pages 422–435. Springer, 1980.
- [21] C. Reutenauer.  $\mathbb{N}$ -rationality of zeta functions. *Adv. in Appl. Math.*, 18(1):1–17, 1997.

UNIVERSITÉ PARIS-EST, LABORATOIRE D'INFORMATIQUE GASPARD-MONGE, UMR 8049  
CNRS

*E-mail address:* `beal@u-pem.fr`

UNIVERSITÉ PARIS-EST, LABORATOIRE D'ALGORITHMIQUE, COMPLEXITÉ ET LOGIQUE

*E-mail address:* `catalin.dima@u-pec.fr`